8-2019

A semi-automated algorithm to quantify similarity between outsole impressions using SURF

Soyoung Park
Iowa State University, sypark@iastate.edu

Alicia L. Carriquiry
Iowa State University, alicia@iastate.edu

A semi-automated algorithm to quantify similarity between outsole impressions using SURF

Disciplines

Forensic Science and Technology

Comments

Posted with permission of CSAFE.

Similarity between outsole impressions using SURF

Soyoung Park, Alicia Carriquiry

csafe

Center for Statistics and Applications in Forensic Evidence

It takes a village

Collaborators:

- Dr. Hari Iyer, NIST
- Sarena Wiesner, Yaron Shor, Israeli Police
- Dr. Guillermo Basulto-Elias, Mr. James Kruse, CSAFE
- Ms. Lesley Hammer, Hammer Forensics
- Dr. Eric Hare, Omni Analytics
- A small army of super smart undergraduates, Iowa State University.

Goals of presentation

- Introduce an objective method to quantify the similarity between two outsole impressions.
- Show that algorithm is accurate and reliable even when outsoles share class characteristics and degree of wear.
- Show that algorithm is robust even when one image is degraded and partially observed.

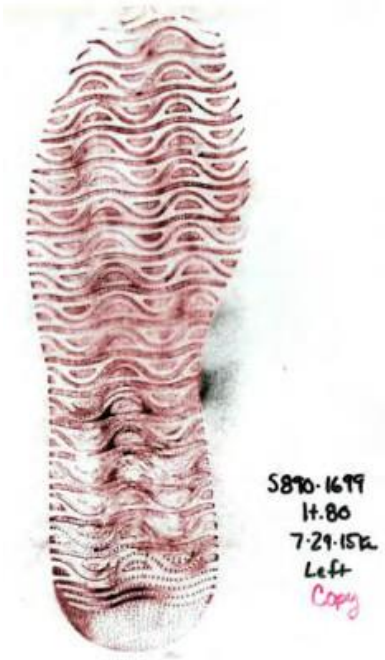
The forensic question

Could the shoe on the left be the source of the impression on the right?

Q = questioned

K = known

Challenging!



The state of the art

- At present, practitioners rely on training and experience to:
  - Identify features of interest that can be used to compare outsoles.
  - Subjectively determine whether outsoles are “similar enough” to suggest same source.
- Two implicit assessments:
  - How similar are the outsoles?
  - How probative is observed degree of similarity.

Can we do better?

- Develop a (semi-)automated algorithm that:
  - Identifies features of interest in outsoles.
  - Compares their values across the two impressions.
  - Computes a *similarity score.*
- Ideally, the similarity score is:
  - Large, when impressions are made by the same shoe.
  - Small, otherwise.

Rich literature

- Bodziak, WJ. 2000. CRC Press.
- Stone, RS. 2006. *J For Ident.*
- Patil, PM., Kulkarni, JV. 2009. *Pattern Recognition.*
- Algarni, G., Hamiani, M. 2008. *For Sci Int.*
- Kortylewski, A. et al. 2014. *Asian Conf Comp Vision.*
- Gwo, CY, Wei, CH. 2016. *Sci & Just.*
- Alizadeh, S., Kose, C. 2017. *For Sci Int.*
- Bodziak, WJ. 2017. CRC Press.
- Richetelli, N., et al. 2017. *For Sci Int.*
- Park, S., Carriquiry, A. 2019. Revision requested, *JCGS.*
- *MANY MORE*

Data

- CSAFE built a database with 160 pairs of shoes.
- Two models: Nike Airflow and Adidas Seeley.
- Four sizes: 8, 8.5, 10, 10.5.
- Participants received a pair of new shoes with a step counter.
- Returned every 6-8 weeks so that shoe outsoles could be photographed and imaged. [Everos 2D scanner.]
- For each shoe: 4 sets of replicated measurements over a 6-month period.
- **To download: www.forensicstats.org/data**.

Outsole photos

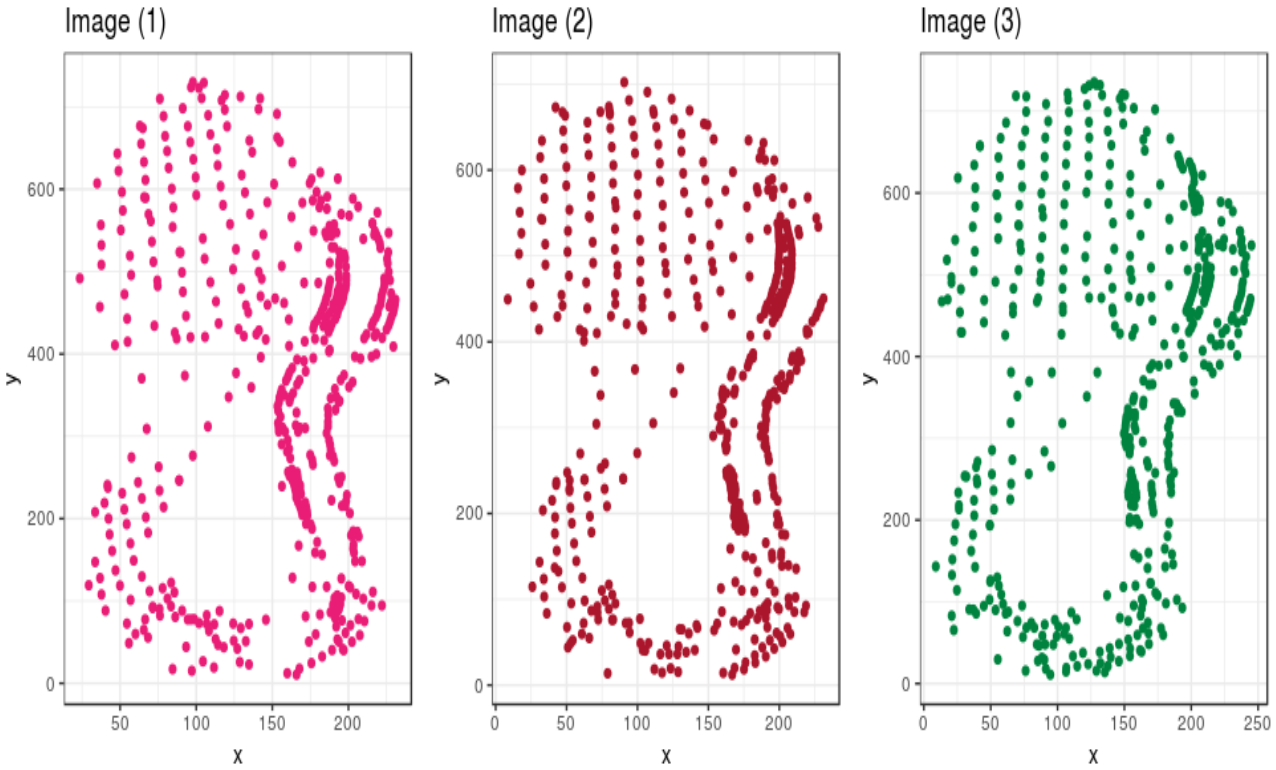
All Nike (Adidas) shoes share class characteristics and degree of wear.

Algorithm I

- We started out with 2D images of 60 pairs of Nike shoe outsoles with about 6 months of wear.
- Each shoe was imaged 4 times.
- Data: The x-y coordinates of the pixels that form the image.
- Each image has thousands of pixels.
- First step: Identify ***interesting points*** in each image.

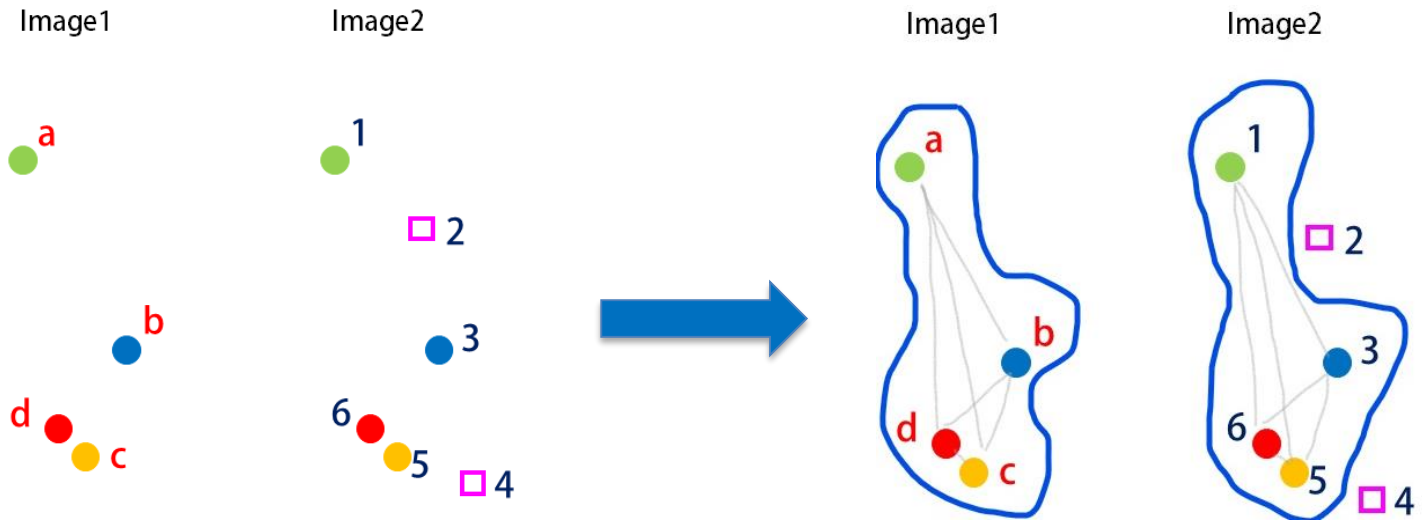
Algorithm II - SURF

- We used 500 SURF (Speeded-Up Robust Features, Bay et al. , 2006) as the points of interest.



Algorithm III -Alignment

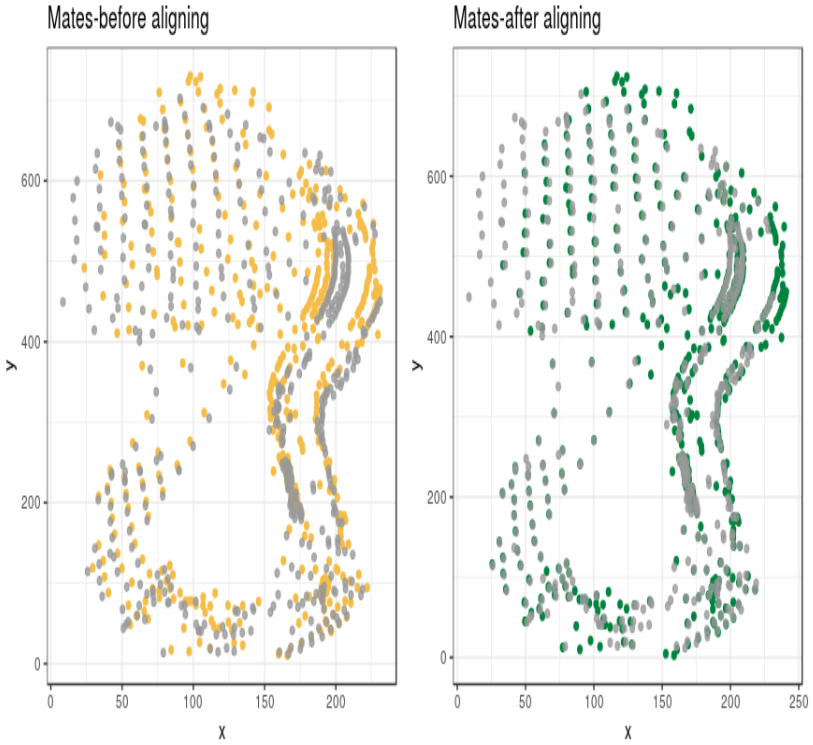
- To align two images Q and K we find the ***maximum clique:*** set of points in Q and K that have the same geometric arrangement.
- The MC is invariant to rotation and translation.



Algorithm IV - Alignment

- Once the MC_Q and the MC_K have been found, the algorithm computes the rotation angle and translation matrix needed to overlay MC_Q onto MC_K.
- These estimates are then used to align all other points on the images.
- Computationally expensive but effective.

Aligned images



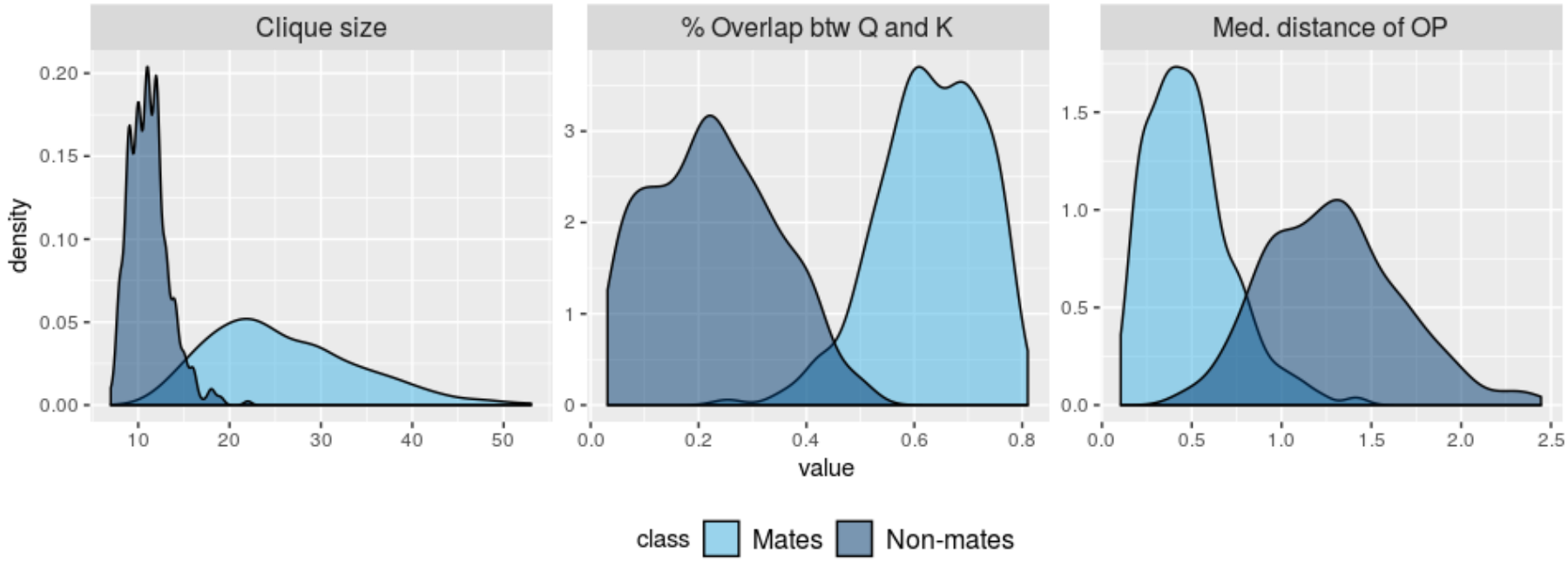
Features

- Size of the MC
- % overlapping points: two points overlap when their distance is 2 pixels or less.
- Median Euclidean distance between overlapping points.
- 501 mated Nike pairs, and 420 non-mated pairs.

Class	Clique size	Rotation angle	% Overlap	Median distance
Mates	18	2.11	56.5%	0.78
Non-mates	9	6.43	12.1%	1.39

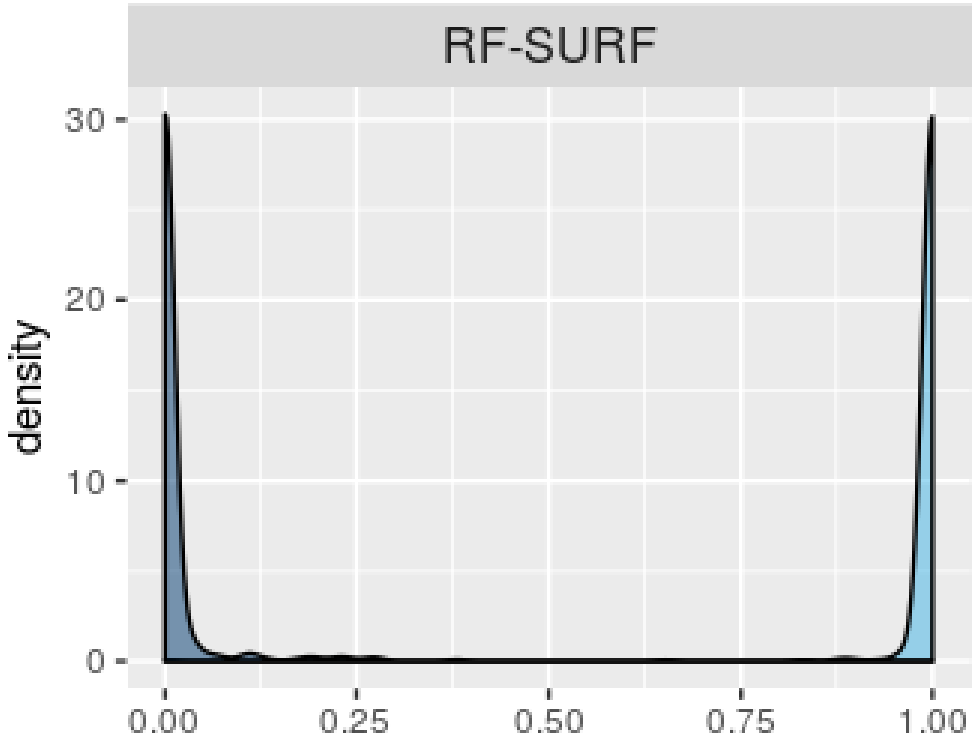
Feature values for M and NM pairs

- Can we tell which pairs of images were made by the same or by different shoes?



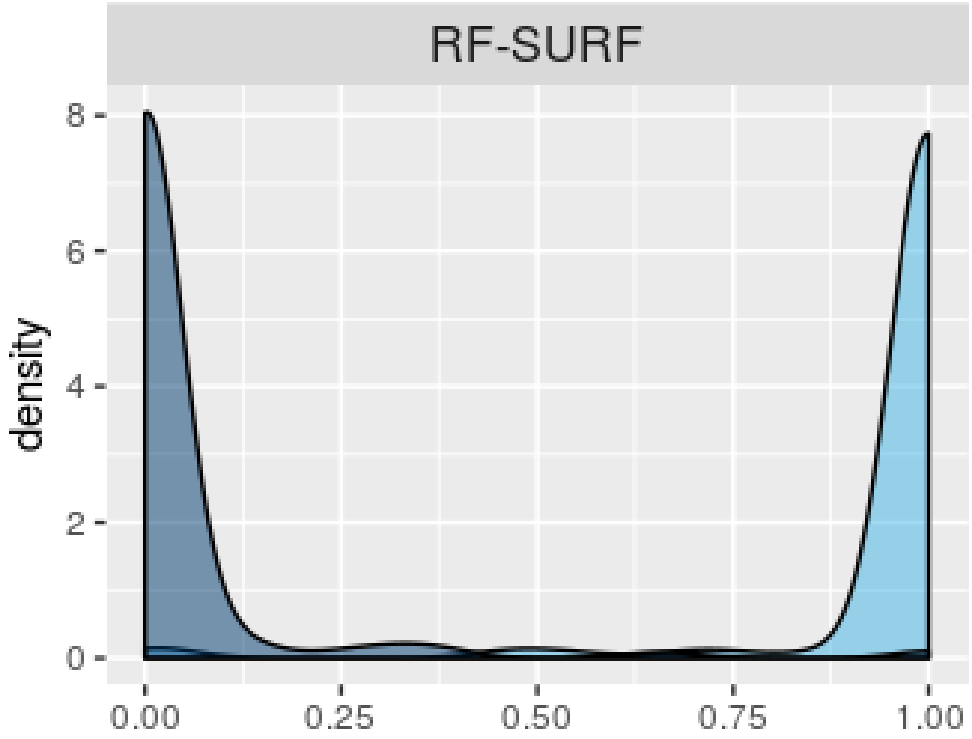
Combining features via RF

- We can combine features into a single ***similarity score.***
- The score is HIGH when images come from the same shoe.
- Good discrimination!



When used on a different set of shoes…

- Does the classifier work on a different set of images?
- Yes!
- **False positive rate is 2.2%.**
- **False negative rate is 1.9%.**
- **Overall accuracy: 97.8%.**



BUT….

- We developed and tested the algorithm using only Nike shoes.
  - **Will the same model be as accurate if we try to classify images from other brands?**
- Both the Q and the K images are of high quality.
  - **Will the performance of the algorithm suffer when Q is degraded or partially observed?**

Different pattern

- The random forest model was trained and tested on Nike Airflow shoes.
- How will the algorithm work if we present it with images of Adidas Steeley shoes?

Prediction	Images are from same shoe	Images are from different shoes
Predict same shoe	198	0
Predict different shoe	54	210
Total	252	210

Different pattern

- If we retrain the random forest using a mix of Nike and Adidas shoes, we can improve performance.

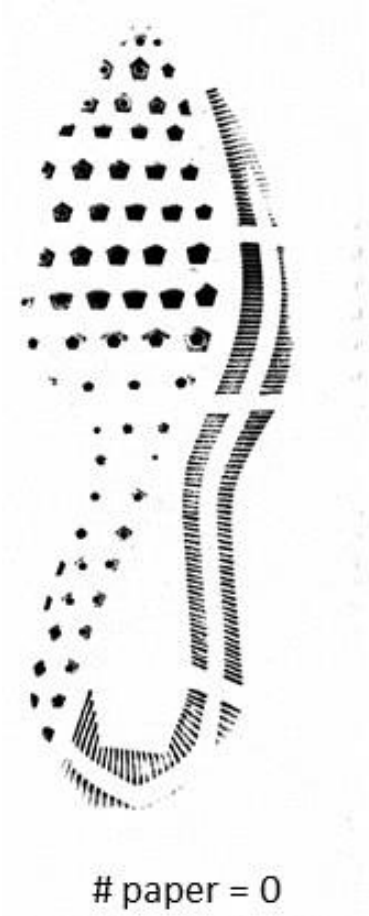
Prediction	Images are from same shoe	Images are from different shoes
Predict same shoe	277	8
Predict different shoe	11	196
Total	288	204

- Need to explore whether it is better to have a single model for all shoes or brand-level models.

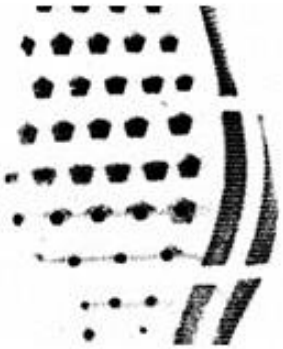
Degraded images II

- Often, the latent print at the crime scene is blurry or partially observed.
- To test the performance of the algorithm, we degraded the Q image by interposing 2, 4, 6, 8 or 10 sheets of paper between the outsole and the scanner.
- We also deleted about half of the image.
- The K image was still of high quality.

Degraded images III



paper = 0



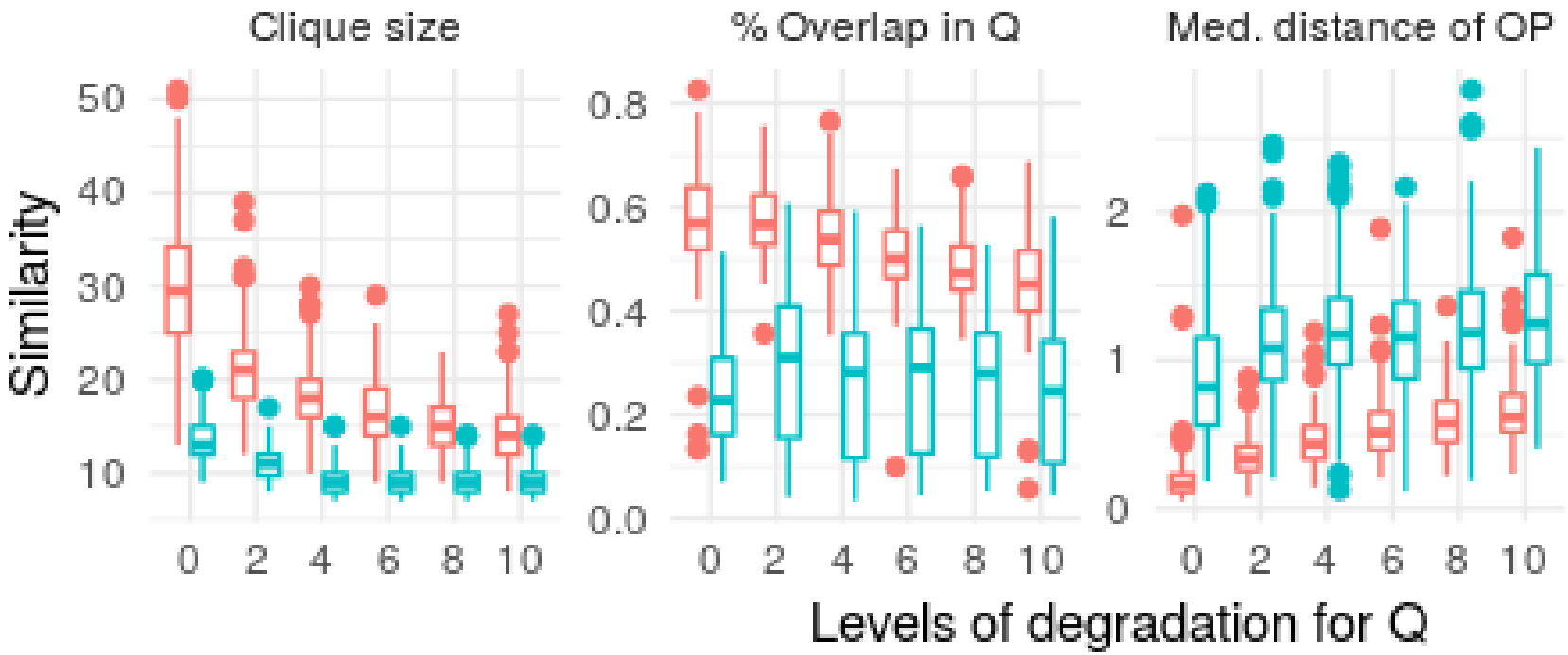
paper = 2

paper = 6

paper = 10

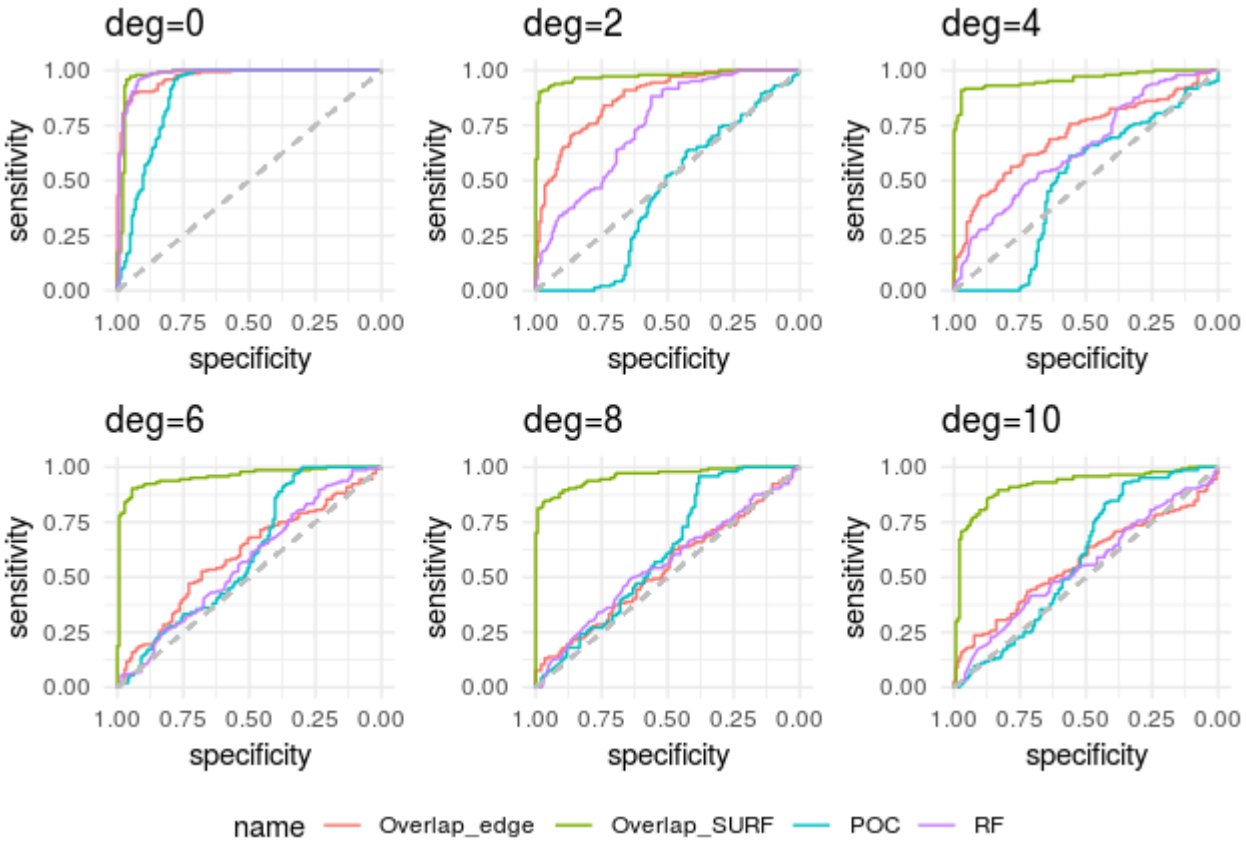
Performance

- Red: mates, Blue: non-mates



Performance II

X-axis: True negative rate, Y-axis: True positive rate



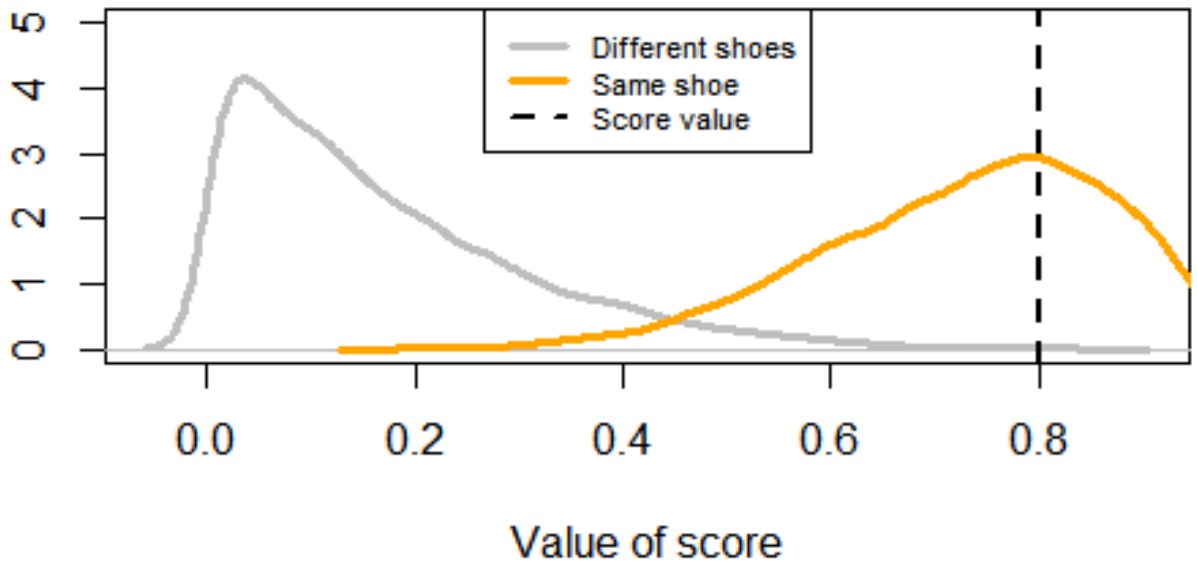
% overlapping SURF

- Classifier that relies only on the % of SURFs in Q and K that overlap appears to be robust to this particular type of image degradation.
- Results for most degraded Q (10 sheets of paper between shoe and scanner):
  - True positive rate: 91% -- Correctly conclude same shoe
  - True negative rate: 97% -- Correctly conclude different shoe
  - **False positive rate: 3% -- Incorrectly conclude same shoe**
  - False negative rate: 9% -- Incorrectly conclude diff shoe

Probative value of score?

- Looks like we might be able to construct a similarity score with good performance.
- Suppose that in the course of an investigation, we get a pair of prints to compare.
- We compute the similarity score and say we get a value equal to 0.8.
- What does the value tell us? Does it suggest same or different shoe?

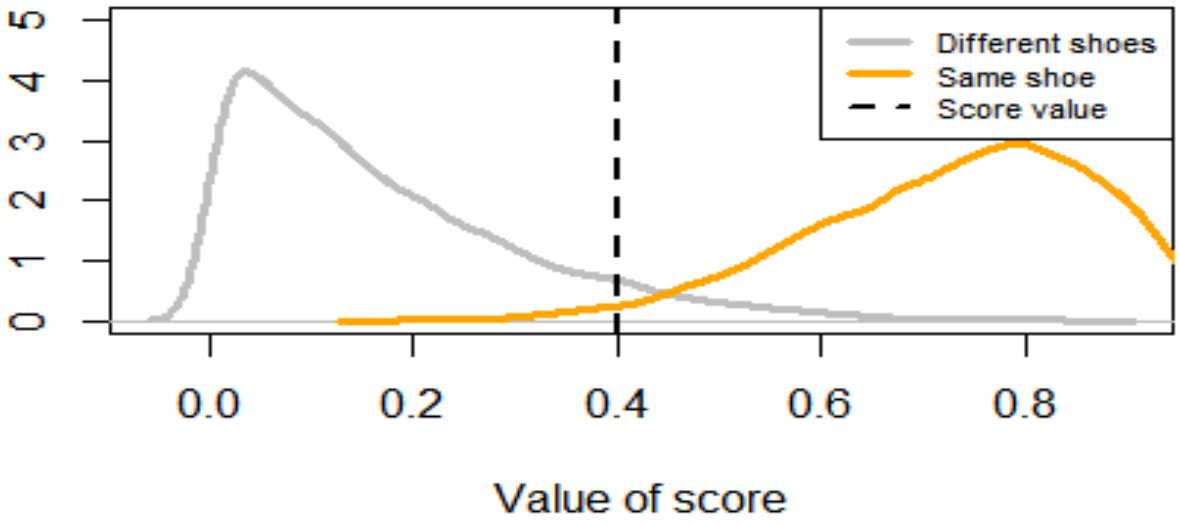
Score of 0.8



Score-based LR = 2.8 / 0.008 = 357.

It is 357 times more likely to observe a score of 0.8 when images are from the same shoe than when they are not.

Score of 0.4



Score-based LR = 0.4 / 1.1 = 0.36.

It is 2.7 times more likely to observe a score of 0.4 when images are from different shoes.

In summary…

- Automatically finding interesting points in an outsole using SURF appears promising.
- When both impressions are of reasonably good quality, an algorithm based on three features has excellent performance, at least for Nike Airflows and Adidas Steeley.
- When Q is degraded, features no longer serve to predict whether Q and K were made by same shoe.
- **Exception is % points that overlap; this feature is robust on degraded and partially observed images.**

Open questions

- Will algorithms continue to perform well when:
  - We include other brands of shoes (we think yes).
  - Q is degraded in some other way (we need more research).
- What data do we need in order to calculate probative value via SLRs?
- Black-box study: compare the outcome of the automated method to the scores produced by a trained examiner.

THANKS!

- Please find me at:

Alicia@iastate.edu

- Our website:

www.forensicstats.org

- Webinar about pattern evidence research at CSAFE on August 20, at 12:00 CT.